Supporting Information

Peter et al. 10.1073/pnas.1207852109

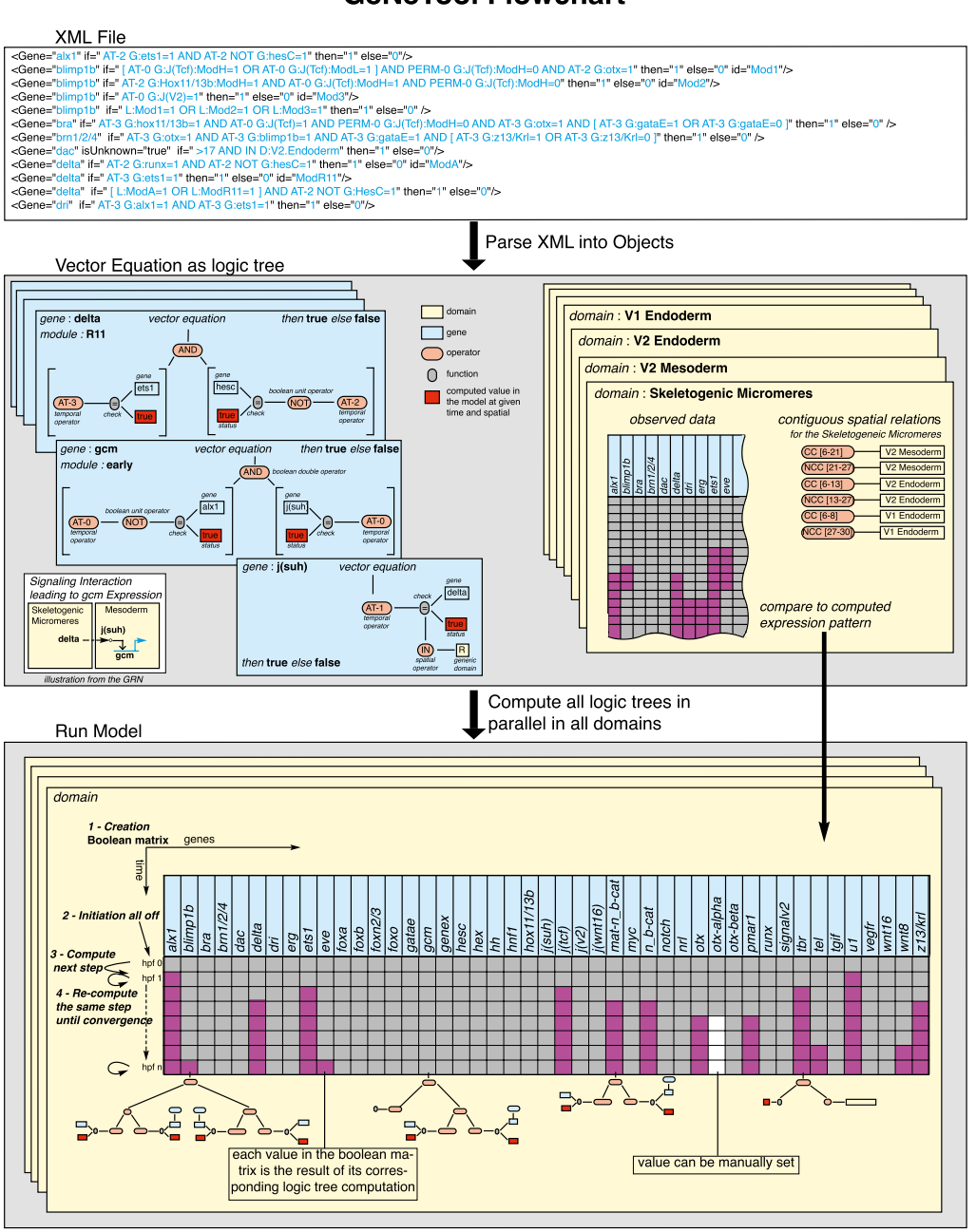


Fig. S1. Flowchart for computation of spatial expression patterns in time by the GeNeTool software. In the box at the top a few of the vector equations in Fig. S3 are reproduced, to indicate their input into the calculation (black arrow, "Parse XML into objects"). The three examples given in the blue panel immediately below show the three logical operations spelled out by the vector equations for the three components of the simplified network circuit shown at lower left. In the logic tree diagrams, "check" indicates that the program is required to ascertain the presence of the indicated inputs, either "on" (true) indicated here by red, or "off," indicated by an open box, the number of hours earlier shown in the associated "at" statement. The logic operators "and" and "not" are called in as indicated; other operators are "in," stating the relevant location of the "R" spatial domain (Fig. S9 shows all operations used in GeNeTool). The four embryonic spatial domains considered in this work are shown on yellow backgrounds (*Right*). In addition, a sample of the observed Boolean matrix of spatial gene expression patterns is portrayed: each horizontal row pertains to a particular time of observation, the next row to the next time. In our study, the observed expression was monitored by whole-mount in situ hybridization at 3-h time intervals. *Right*: Example of how spatial contiguity of cells of the domain shown on the top of the stack, the skeletogenic domain, is coded with respect to cells of other domains: e.g., at 6 to 21 h, these cells are contiguous (CC) to cells of veg2 mesoderm; at 21 to 27 h, they are not contiguous (NCC) to the veg2 mesoderm. *Lower*: Operation of a GeNeTool computation. At each hour, the state

Legend continued on following page

of expression of each gene in the system (top row) is predicted according to its logic tree computation, four of which are indicated below. That is, as indicated by the curved arrows, the inputs to the computation at each location (square) in the gene/time matrix is computed from the logic relations mandated in the appropriate vector equation, using the inputs computed at the locations required in the equation earlier. The results are reported as predicted to be expressed (colored square) or predicted to be not expressed (gray square).

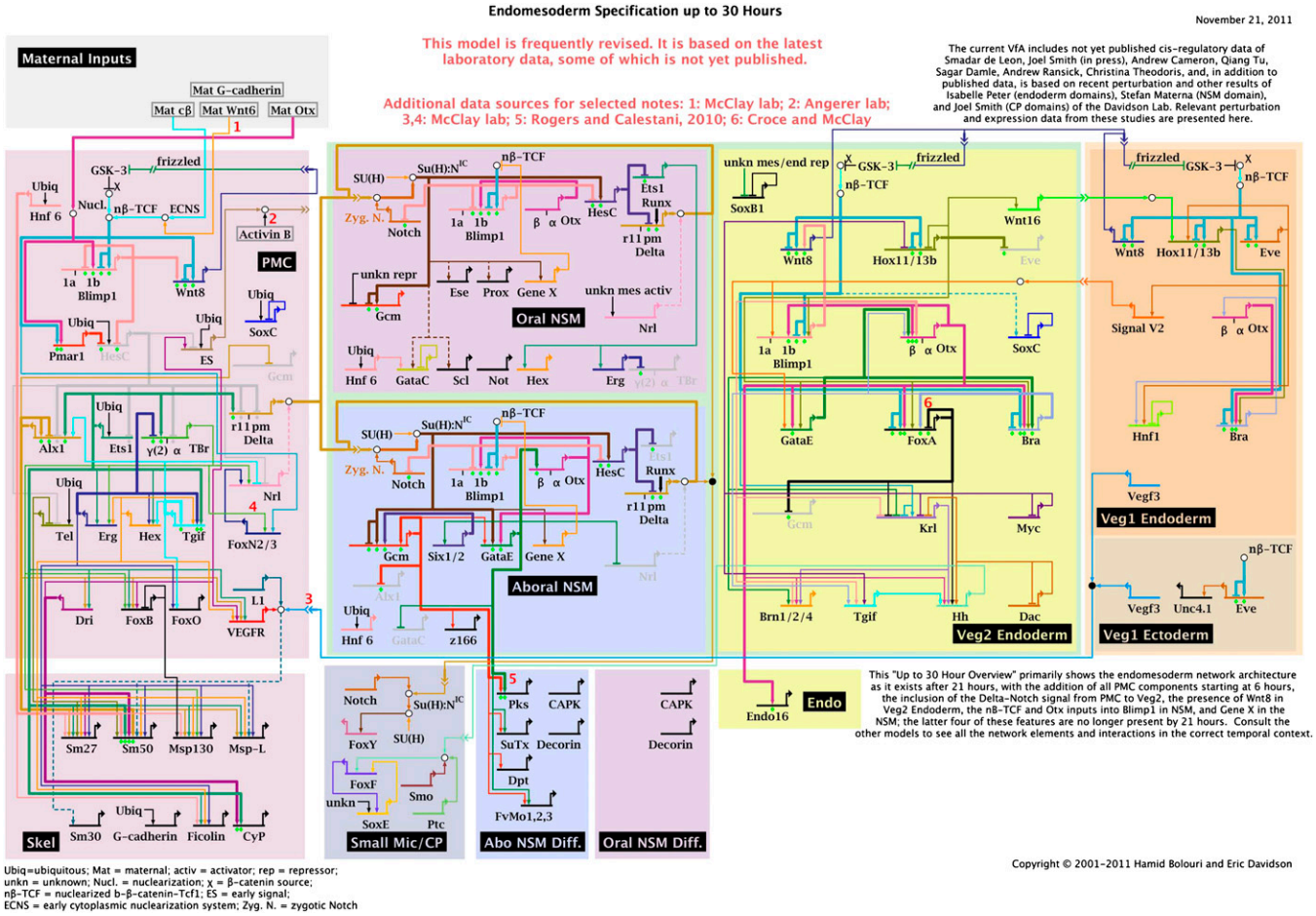


Fig. S2. Current version of endomesodermal gene regulatory network (GRN) model, 18 to 30 h. The model is continuously updated and is available online (http://sugp.caltech.edu/endomes/#BioTapestryViewer), together with underlying expression and perturbation data upon which it is based, and detailed symbolism. At each node, the inputs and outputs of the gene indicated can be followed by color coding. Functional linkages directly substantiated by *cis*-regulatory analysis are shown in heavier lines. GRN model is presented in the BioTapestry software platform (1).

1. Longabaugh WJ, Davidson EH, Bolouri H (2009) Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim Biophys Acta* 1789: 363–374.

Gene		Vector Equation
alx1		if AT-2 *ets1*=1 AND AT-2 NOT *hesc*=1 then=1 else=0
blimp1b		if Mod1=1 OR Mod2=1 OR Mod3=1 then=1 else=0
	Mod1	if [AT-0 *j(tcf)*:ModH=1 OR AT-0 *j(tcf)*:ModL=1] AND PERM-0 *j(tcf)*:ModH=0 AND AT-2 *otx*=1 then=1 else=0
	Mod2	if AT-2 *hox11/13b*:ModH=1 AND AT-0 *j(tcf)*:ModH=1 AND PERM-0 *j(tcf)*:ModH=0 then=1 else=0
	Mod3	if AT-0 *j(v2)*=1 then=1 else=0
bra		if AT-3 *hox11/13b*=1 AND AT-0 *j(tcf)*=1 AND PERM-0 *j(tcf)*:ModH=0 AND AT-3 *otx*=1 AND [AT-3 *gatae*=1 OR AT-3 *gatae*=0] then=1 else=0
brn1/2/4		if AT-3 *otx*=1 AND AT-3 *blimp1b*=1 AND AT-3 *gatae*=1 AND [AT-3 *z13/krl*=1 OR AT-3 *z13/krl*=0] then=1 else=0
dac		if >17 AND IN V2 Endoderm then=1 else=0
delta		if [ModA=1 OR ModR11=1] AND AT-2 NOT *hesc*=1 then=1 else=0
	ModA	if AT-2 *runx*=1 AND AT-2 NOT *hesc*=1 then=1 else=0
	ModR11	if AT-3 *ets1*=1 then=1 else=0
dri		if AT-3 *alx1*=1 AND AT-3 *ets1*=1 then=1 else=0
erg		if AT-3 *ets1*=1 AND AT-3 *tbr*=1 OR AT-3 *hex*=1 then=1 else=0
ets1		if AT-0 *u1*=1 AND AT-2 NOT *hesc*=1 then=1 else=0
eve		if AT-2 *j(tcf)*=1 AND PERM-0 *j(tcf)*:ModH=0 AND PERM-3 [*hox11/13b*:ModH=1 AND *eve*=1] then=1 else=0
foxa		if Mod1=1 OR Mod2=1 OR Mod3=1 then=1 else=0
	Mod1	if [AT-0 *j(tcf)*:ModH=1 OR AT-0 *j(suh)*=1] AND PERM-0 *j(tcf)*:ModH=0 then=1 else=0
	Mod2	if AT-0 *j(tcf)*:ModH=1 AND AT-3 *hox11/13b*:ModH=1 AND AT-3 *otx*=1 AND [AT-3 *bra*=1 OR AT-3 *bra*=0] AND PERM-0 *j(tcf)*:ModH=0 then=1 else=0
	Mod3	if >23 AND IN V2 Endoderm then=1 else=0
foxb		if AT-3 *alx1*=1 AND AT-3 *dri*=1 AND AT-3 *ets1*=1 AND AT-3 *tbr*=1 then=1 else=0
foxn2/3		if AT-3 *tbr*=1 then=1 else=0
foxo		if AT-3 *tgif*=1 AND AT-3 *erg*=1 then=1 else=0
gatae		if [AT-3 *otx*=1 AND AT-0 *j(suh)*=1 AND AT-3 *gcm*=1] OR [AFTER-3 *hox11/13b*:ModH=1 AND AT-0 *j(v2)*=1 AND AT-3 *otx*=1] then=1 else=0
gcm		if ModE=1 OR ModG=1 then=1 else=0
	ModE	if AT-0 *j(suh)*=1 AND NOT *alx1*=1 AND PERM-0 *j(suh)*=0 then=1 else=0
	ModG	if AT-2 *gcm*=1 AND AT-2 NOT *foxa*=1 then=1 else=0
genex		if >15 AND AT-0 *j(suh)*=1 then=1 else=0
hesc		if [AT-0 *u1*=1 OR AT-0 *j(suh)*=1] AND PERM-0 *pmar1*=1 then=1 else=0
hex		if AT-2 *erg*=1 AND [AT-2 *ets1*=1 OR AT-2 *tgif*=1] then=1 else=0
hh		if AT-3 *dac*=1 AND AT-3 *foxa*=1 AND AT-3 *tgif*=1 AND AT-3 *otx*=1 AND [AT-3 *z13/krl*=1 OR AT-3 *z13/krl*=0] then=1 else=0
hnf1		if AT-2 *bra*=1 AND AT-2 *eve*=1 then=1 else=0
hox11/13b		if ModH=1 OR ModW=1 then=1 else=0
	ModH	if AT-0 *j(tcf)*:ModH=1 AND NOT [AT-3 *j(wnt16)*=1 AND AT-3 *hox11/13b*:ModH=1] AND

Fig. S3. (Continued)

		PERM-0 *hox11/13b*:ModH=0 AND [AT-2 *myc*=1 OR AT-2 *myc*=0] AND PERM-0 *j(tcf)*:ModH=0 then=1 else=0
	ModW	if AT-2 *eve*=1 AND AT-0 *j(wnt16)*=1 then=1 else=0
j(suh)		if AT-1 *delta*=1 IN CC R then=1 else=0
j(tcf)		if ModH=1 OR ModL=1 OR ModV1=1 then=1 else=0
	ModH	if AT-0 *n_b-cat*:ModH=1 OR AT-1 *n_b-cat*:ModH=1 then=1 else=0
	ModL	if AT-0 *n_b-cat*:ModL=1 OR AT-1 *n_b-cat*:ModL=1 then=1 else=0
	ModV1	if AT-0 *n_b-cat*:ModV1=1 OR AT-1 *n_b-cat*:ModV1=1 then=1 else=0
j(v2)		if AT-0 *signalv2*=1 IN CC R then=1 else=0
j(wnt16)		if AT-3 *wnt16*=1 IN CC R OR AT-3 *wnt16*=1 IN R then=1 else=0
mat-n_b-cat		if ModSkel=1 OR ModV2E=1 OR ModV1E=1 OR ModV2M=1 then=1 else=0
	ModSkel	if >4 AND <10 AND IN Skeletogenic Micromere then=1 else=0
	ModV2E	if >5 AND <13 AND IN V2 Endoderm then=1 else=0
	ModV1E	if IN V1 Endoderm AND >5 AND <10 then=1 else=0
	ModV2M	if IN V2 Mesoderm AND >5 AND <13 then=1 else=0
myc		if ModV2M=1 OR ModV2E=1 then=1 else=0
	ModV2M	if >14 AND IN V2 Mesoderm then=1 else=0
	ModV2E	if >14 AND IN V2 Endoderm then=1 else=0
n_b-cat		if ModH=1 OR ModL=1 OR ModV1E=1 then=1 else=0
	ModL	if AT-0 *mat-n_b-cat*=1 then=1 else=0
	ModV1E	if IN V1 Endoderm AND >9 AND <31 AND NOT *genex*=1 then=1 else=0
	ModH	if [IN V2 Endoderm OR IN V2 Mesoderm] AND >9 AND <23 AND [AT-0 *wnt8*=1 IN CC R OR AT-0 *wnt8*=1 IN R] AND AT-0 NOT *genex*=1 then=1 else=0
notch		if >20 AND IN V2 Endoderm then=1 else=0
nrl		if [AT-2 *foxn2/3*=1 OR AT-2 *tgif*=1] AND AT-2 *tbr*=1 AND AT-2 NOT *hesc*=1 then=1 else=0
otx		if AT-0 *otx-alpha*=1 OR AT-0 *otx-beta*=1 then=1 else=0
otx-alpha		if ModSkel=1 OR ModV2=1 OR ModV1E=1 then=1 else=0
	ModSkel	if >4 AND <24 AND IN Skeletogenic Micromere then=1 else=0
	ModV2	if >7 AND <27 AND [IN V2 Endoderm OR IN V2 Mesoderm] then=1 else=0
	ModV1E	if IN V1 Endoderm AND >7 AND <27 then=1 else=0
otx-beta		if [AT-3 *gatae*=1 AND AT-3 *otx-alpha*=1] OR [AT-3 *blimp1b*=1 OR AT-3 *bra*=1] then=1 else=0
pmar1		if [AT-0 *j(tcf)*:ModH=1 OR AT-3 *j(tcf)*:ModL=1 OR AT-0 *j(tcf)*:ModL=1] AND PERM-3 *j(tcf)*:ModH=0 AND AT-0 *otx*=1 AND AT-2 NOT *hesc*=1 then=1else=0
runx		
signalv2		if IN V1Endoderm AND AT-3 *eve*=1 AND AT-3 *hox11/13b*=1 then=1 else=0
tbr		if [ModA=1 OR ModG2=1] AND AT-3 NOT *hesc*=1 then=1 else=0
	ModA	if AT-3 *ets1*=1 AND AT-3 NOT *hesc*=1 then=1 else=0
	ModG2	if AT-3 *ets1*=1 AND AT-3 NOT *erg*=1 then=1 else=0
tel		if AT-3 *u1*=1 AND AT-3 NOT *hesc*=1 then=1 else=0
tgif		if ModE=1 OR ModL=1 then=1 else=0
	ModE	if AT-3 *ets1*=1 AND AT-3 *hex*=1 AND AT-3 *erg*=1 then=1 else=0
	ModL	if [AT-3 *blimp1b*=1 AND AT-3 *myc*=1] AND [AT-3 *bra*=0 OR AT-3 *bra*=1] then=1 else=0
u1		if AFTER-1 *pmar1*=1 then=1 else=0

vegfr		if AT-3 *alx1*=1 AND AT-3 *dri*=1 AND AT-3 *hex*=1 AND AT-3 *ets1*=1 then=1 else=0
wnt16		if AT-3 *hox11/13b*=1 AND AT-3 *bra*=1 then=1 else=0
wnt8		if AT-1 *j(tcf)*=1 AND PERM-0 *j(tcf)*:ModH=0 AND [AT-2 *blimp1b*=1 OR AT-2 *blimp1b*=0] then=1 else=0
z13/krl		if [AT-0 *j(tcf)*:ModH=1 OR AT-0 *j(tcf)*:ModL=1] AND PERM-0 *j(tcf)*:ModH=0 AND NOT [AT-2 *foxa*=1 AND AT-2 *gatae*=1 AND AT-2 *dac*=1 AND AT-2 *blimp1b*=1 AND AT-2 *myc*=1] then=1 else=0

Fig. S3. Complete set of vector equations used to power the computational model. As described in text, the inputs and logic functions are derived primarily from the GRN models; Fig. 1 provides information on epistemology.

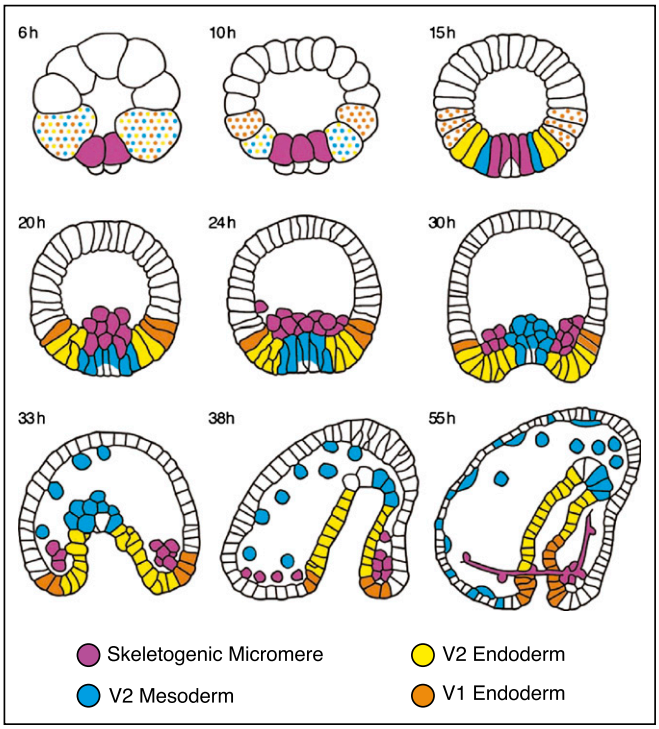
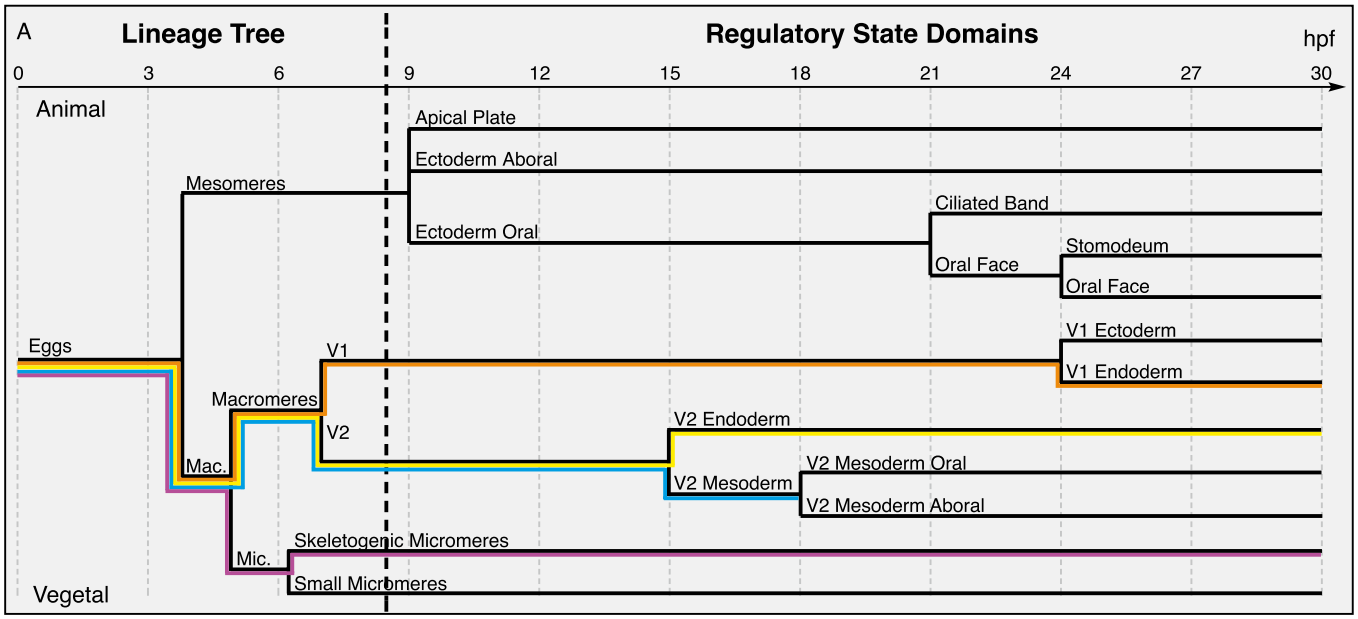


Fig. S4. Standard diagram of sea urchin embryogenesis as in *Strongylocentrotus purpuratus* at 15 °C (1). Endomesodermal cell fates and lineages are color coded as in the key and as throughout the study. Embryos are viewed from the side, with vegetal pole at bottom and oral ectoderm to right.

1. Davidson EH, Cameron RA, Ransick A (1998) Specification of cell fate in the sea urchin embryo: Summary and some proposed mechanisms. *Development* 125:3269–3290.



B

	Small Micromeres	Skeletogenic Micromeres	V2 Mesoderm	V2 Endoderm	V1 Endoderm	V1 Ectoderm	Oral Face	Stomodeum	Ciliated Band	Ectoderm Aboral	Apical Plate
Small Micromeres		CC [6-21]	CC [21-30]								
Skeletogenic Micromeres			CC [6-21] NCC [21-27]	CC [6-13] NCC [13-27]	CC [6-8] NCC [27-30]	NCC [27-30]					
V2 Mesoderm				CC [18-30]							
V2 Endoderm					CC [7-30]						
V1 Endoderm						CC [21-30]					
V1 Ectoderm							CC [6-15]	CC [6-15]	CC [6-30]	CC [6-30]	
Oral Face								CC [18-30]	CC [18-30]		
Stomodeum											
Ciliated Band										CC [9-30]	CC [9-30]
Ectoderm Aboral											CC [9-30]
Apical Plate											

Fig. S5. Digital representation of spatial relationships through developmental time for *S. purpuratus*. (*A*) Developmental ancestry of the domains. *Left*: Canonical cell lineage of embryo; each vertical branch represents a cleavage. To the right of the dashed line, the regulatory state and cell fate domains and their derivatives are shown. (*B*) The spatial relationships between the given domain and the listed other territories are given over the periods when they obtain; CC, immediately contiguous; NCC, not contiguous.

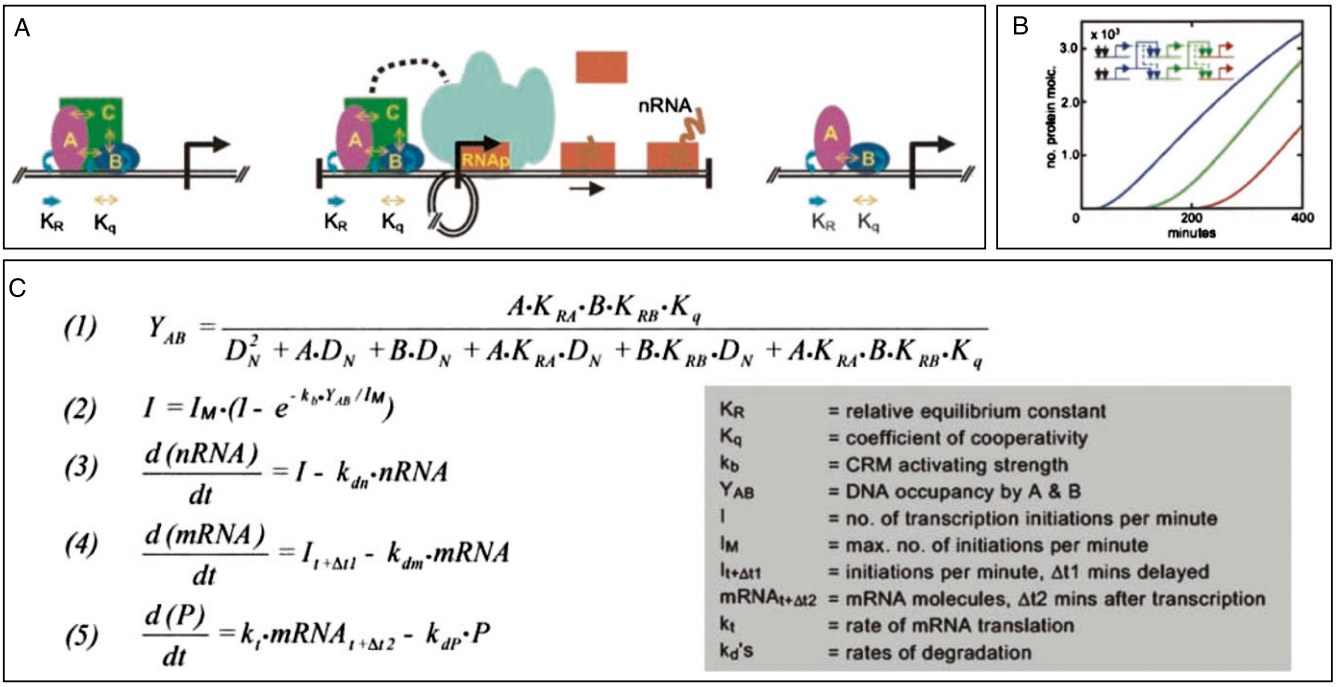


Fig. S6. Model for effect of *cis*-regulatory occupancy on transcriptional output. (*A*) Illustration representing *cis*-regulatory occupancy and transcription complex. (*Left*) CRM requiring three transcription factors (marked "A," "B," and "C"). All factors interact energetically with one another when brought into proximity, as well as with the DNA, i.e., bind cooperatively. Relative equilibrium constants (K_R) for interaction with DNA are marked in blue, and cooperativity constants (K_q) are marked in yellow. (*Center*) DNA looping results in proximity of fully loaded CRM to the basal transcription apparatus (teal), activation thereof, and recruitment of polymerase (orange); transcription ensues. (*Right*) Two-factor CRM, for which model equations are constructed. (*B*) Computed simulation of step time for a cascade consisting of three steps as shown in the diagram. (*C*) Mathematical model capturing functions diagrammed in *A*. (For derivations, see ref. 1.) Eq. **1** defines double occupancy of the CRM. Eq. **2** defines the relation between occupancy and initiation rate. Eqs. **3**–**5** define the kinetics of nRNA, mRNA, and protein ("P") output. Terms are defined in the box on the right. (Reproduced with permission from ref. 1.)

1. Bolouri H, Davidson EH (2003) Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc Natl Acad Sci USA* 100:9371–9376.

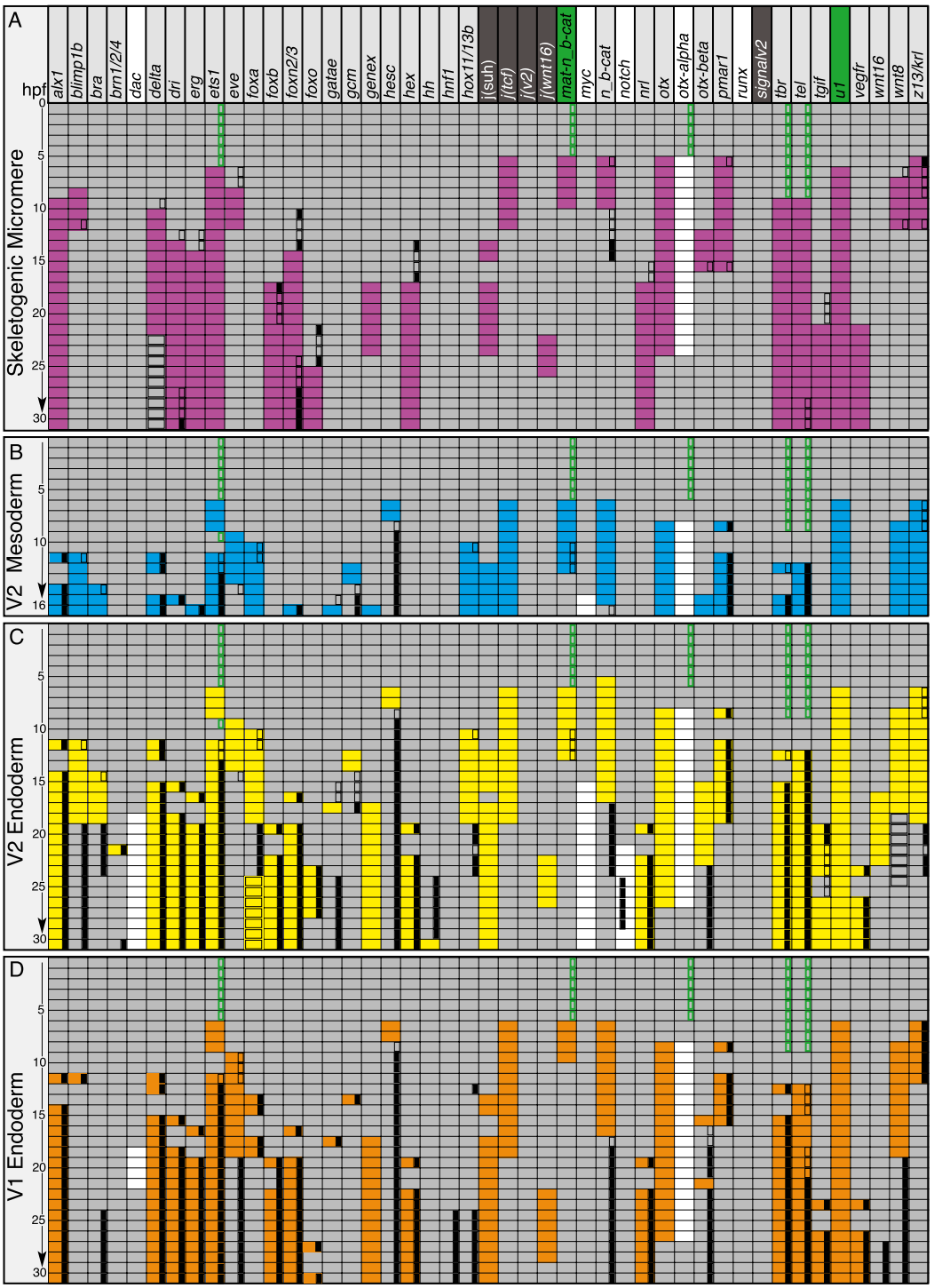


Fig. S7. Effect on model performance of altering the step time to 4 h. Step times were altered in all vector equations from 3 h to 4 h. The form of the figure is the same as that of Fig. 3, showing the four spatial domains (*A–D*). Comparison with Fig. 3 illustrates the dramatic increase in severe deviations from the observed expression data caused by alteration of step time, demonstrating the importance of an accurate estimation of the step time.

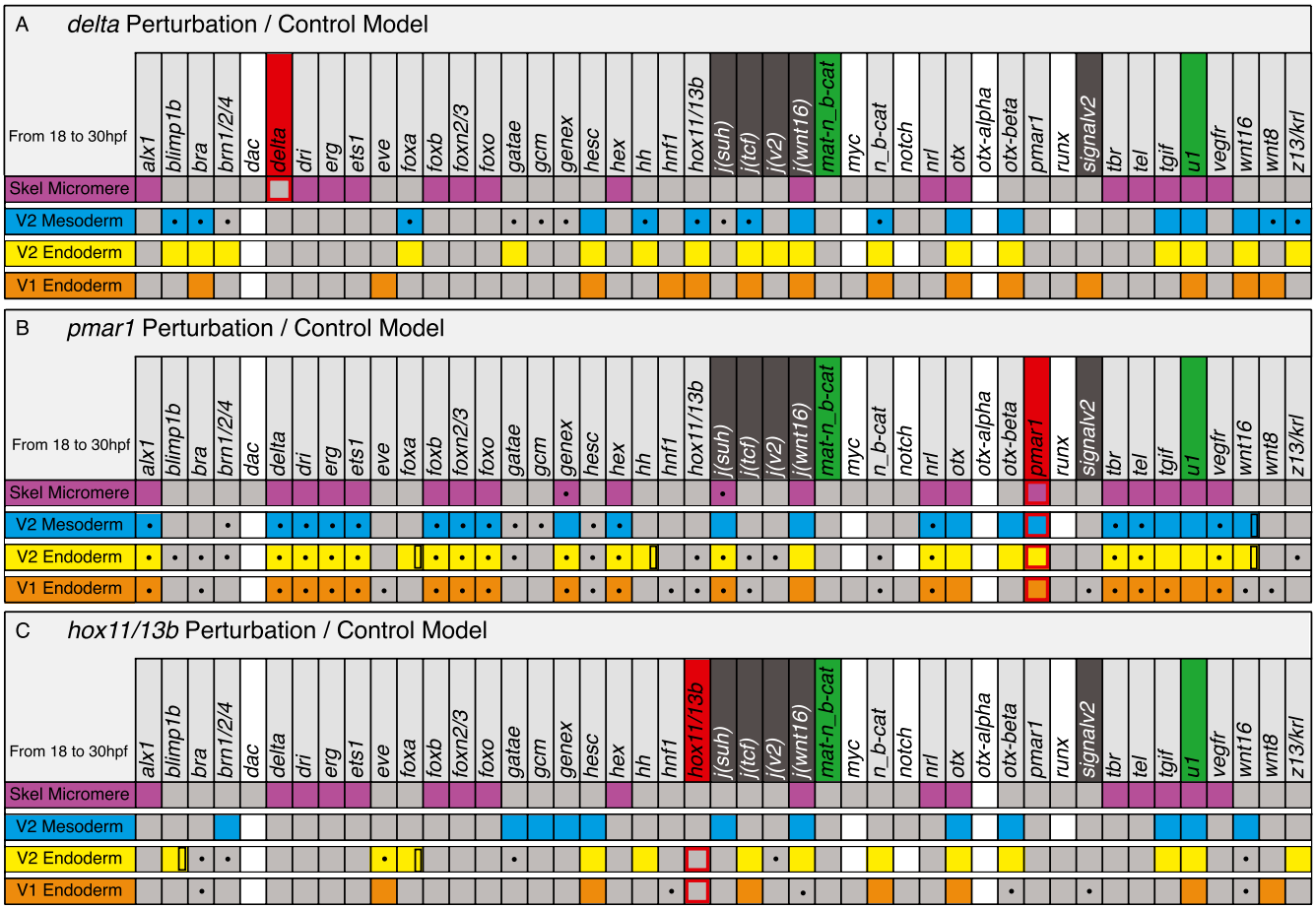


Fig. S8. Detailed computational results of (*A*) *delta*, (*B*) *pmar1*, and (*C*) *hox11/13b* perturbations *in silico*. Presentation is as in Fig. 2. The manually set changes used to mimic the experimental perturbations are indicated in red; red boxes around gray squares mean the gene was manually turned off in the indicated domains; red boxes around colored squares mean the gene was manually turned on in the indicated domains. In cases in which the computational perturbation resulted in a change from off to on, or vice versa, i.e., compared with the control computational model, the field is marked by a black dot; otherwise, there was no change.

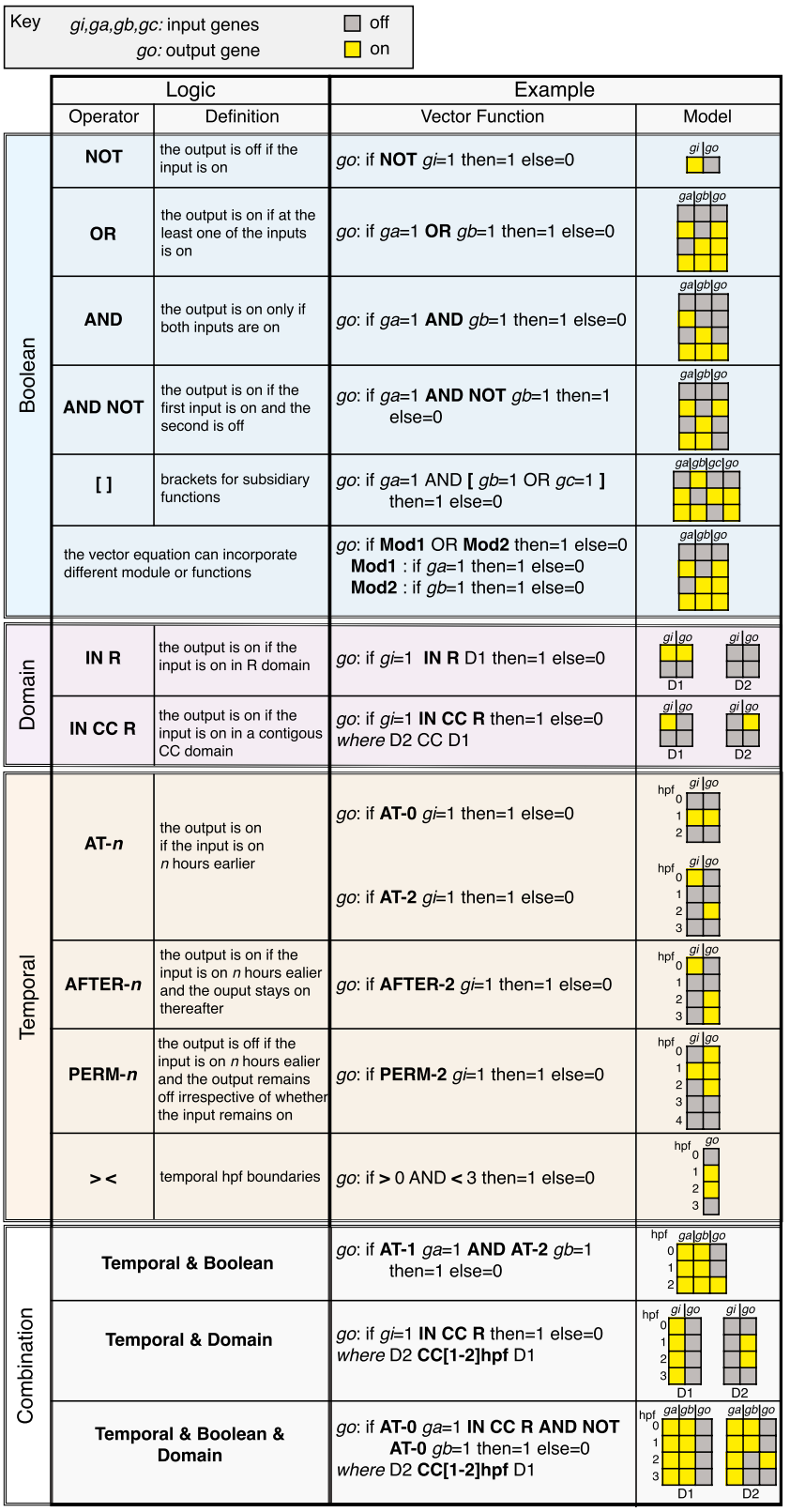


Fig. S9. Operational logic rules for Boolean computation model.